Issues from NSF Review

14 Feb DRAFT

1) There is no clear prioritization of the projects in the proposal to take account, for example, of the possibility of partial funding, or to be used in developing telescope scheduling algorithms. What is your priority in terms of the science?

At the beginning of the planning for the SDSS-II, the ranking
1: Legacy, 2: SEGUE, 3: Supernova, was suggested, which reflected the commitment among the partners to the original goals of the extragalactic survey (Legacy), and the relatively small number of scientists at the participating institutions involved in supernova surveys. Since then, the institutional interest in the Supernova survey has blossomed, and the team has added individuals with direct experience with similar surveys and techniques.

We briefly considered a two-year survey to reduce costs, but we immediately realized that three Spring seasons were needed to accommodate the Legacy and SEGUE spectroscopy in the North Galactic Cap. Running for the necessary three years allows us to include the Supernova survey with little extra cost (see answer to Question 4). The result is that we can undertake three major scientific initiatives; we make the SDSS-II more attractive and have added partners accordingly; and we make good use of the observing systems throughout the year.

As the question suggests, what really matters is how we manage to point the telescope. One needs a set of rules to do this, and a mechanism for making informed decisions that are properly communicated. The rules are as follows, in priority order:

1) If an unobserved Legacy stripe or spectroscopic plate is available in a part of the North Galactic Cap between stripes 10 and 37 that is currently accessible, observe that stripe or plate.

2) September, October, and most of November are allocated to the Supernova survey, where imaging is attempted even in spectroscopic conditions. The right ascension range of the Supernova survey is from 20 h to 4 h; whenever this area is not accessible for at least 1.5 hours at an hour angle of less than 3.25 h, the time is given to SEGUE (e.g., the ends of the nights later in the Fall). Some smaller amount of time may be given to SEGUE in September and October so that SEGUE can obtain stripes and tiles at lower declination that can best be obtained in those months.

3) all other time is given to SEGUE. The choice between imaging and spectroscopy depends on atmospheric conditions (imaging if photometric, good seeing, and dark) and availability of sky. The imaging for a region of sky must proceed the spectroscopy, which is why there is no SEGUE imaging in 2008

For each of the three surveys, we know the total amount of data that needs to be collected to accomplish the science goals, and from this we can estimate the total clock time necessary, given our historical rates of collecting data -- that is, factors for weather and operational efficiency are determined empirically. We have constructed a Baseline Plan for SDSS-II (supplied document) that takes into account these rules and rates, assuming a particular (adjustable) model for allocating time between SEGUE and Legacy. It demonstrates that the total time in a three-year duration is adequate. Because its goal is to understand the full structure of the Milky Way, SEGUE observations are distributed essentially uniformly in right ascension. Thus at any given time, there will be a SEGUE field that is observable, making scheduling much easier.

Any more detailed a priori plan would of course be immediately rendered obsolete by the enormous uncertainties due to weather. Thus, guided by the baseline plan, we make night-to-night and darkrun-to-darkrun decisions that allow us to continuously adapt to weather conditions. In SDSS-I we have had to balance between imaging and spectroscopy, and between observing in the Northern and Southern Galactic Caps. We can build on this experience for scheduling SDSS-II, as we now describe.

We drill separate spectroscopic plates for Legacy and for SEGUE, using the target selection algorithm for the respective program. Nine fiber cartridges are plugged during the day to be available at night, according to a daily evaluation of the spectroscopic needs for that night. We have a plate database that provides priorities based on hour angle, airmass and science goals. Exchanges between fiber cartridges is fast. It is also relatively easy and fast to mount the imaging camera during a night, if the conditions allow imaging.

Between dark runs, Steve Kent (as Head of Survey Coordination, see SDSS-II Management Plan, supplied document) reviews the success of observations from the previous dark run and anticipates what should be accomplished in the next dark run to optimize the efficiency of the overall survey. This process results in a "monthly observing plan," namely a set of instructions about priorities, that is communicated to the Observers. On a given night, the Observers are responsible for the actual selection of observations. They are PhD astronomers and full scientific Participants in the science, they are familiar with the nature of the trade-offs, and they have the expertise to adapt to the prevailing conditions at the mountain. During a dark run, good communications are maintained between Steve Kent and the Observers. Between dark runs, the Project Team Leaders communicate additional information to Steve Kent.

To address more directly the question about partial funding and priorities for the science, our agreement with the Sloan Foundation is that we require funding for the proposed three-year survey. A two-year survey would not be of interest to the Sloan Foundation, nor would it be of interest to the contributing institutions, because it would not achieve our science goals. We have already indicated that dropping the Supernova survey would not result in speeding up the time-to-completion for Legacy and for SEGUE - this is addressed more fully in the answer to Question 4. If we dropped Legacy, it is true that SEGUE could be accomplished without the Spring 2008 time for spectroscopy, but that version of SDSS-II is not what we proposed. The same considerations of course apply if we were to drop SEGUE.

2) Given that there are several groups engaged in SN search programs, why pursue the one you have proposed?

In the "gold" sample of good-quality, published SN Ia lightcurves compiled by Riess, etal (2004), only 6 out of 157 were in the redshift range z = 0.1-0.3. Due to its combination of aperture, field of view, and "natural" scan rate, the SDSS is uniquely suited to obtain high-quality SN Ia lightcurves in this redshift desert, and it will remain so for the next several years. There are several large on-going supernova surveys, e.g., KAIT and the Supernova Factory at low redshift ($z < 0.08$), and ESSENCE and CFHTLS at high redshift (mainly z = 0.3-0.8, though they will have a small number at lower redshift). By 2007, these surveys, along with High-z and SCP, will collectively have accumulated a few hundred SN Ia lightcurves, but the z = 0.1-0.3 region in the SN Hubble diagram will remain substantially underpopulated. There are compelling reasons for targeting this intermediate redshift range with a SN survey comparable in size ($N \sim 200$ lightcurves) to the on-going surveys.

Probing Dark Energy

While multiple lines of evidence indicate that the Universe is accelerating, the current observational constraints on the properties of the dark energy, in particular its equation of state w, are relatively weak. If the dark energy is a cosmological constant ($w = -1$), then it began dominating over non-relativistic matter at about $z = 0.3$. At $z = 0.3$, the difference in distance modulus between cosmological models with different w (keeping other cosmological parameters fixed) is typically about half that at $z = 0.8$. In addition, the locus of constant distance modulus in the plane of w vs. Omega_matter is more nearly orthogonal to the CMB and large-scale structure constraints for SNe at intermediate as opposed to higher redshift. As a result, a SN survey to $z \sim 0.3$ with high statistics and good control of systematics, in combination with these other cosmological measurements, has a surprising amount of power in constraining dark energy. (Note that all of the on-going SN surveys, including SDSS II, will provide only weak constraints on dark energy in the absence of priors from other cosmological measurements.) Moreover, since the line of constant distance modulus rotates in the w-Omega_m plane as the redshift changes, combining the SDSS results with those from, e.g., ESSENCE will lead to tighter constraints than either achieves on its own (as our simulations of these surveys show). Thus, the SDSS SN survey will complement and strengthen the on-going surveys and constitutes a logical next step for ground-based SN studies.

By densely filling in the redshift gap, the SDSS SN survey will enable completion of a continuous Hubble diagram from low- to high-redshift, a measurement of fundamental interest for cosmology. By achieving this, it will be able to uncover (or more likely exclude) any cosmological surprises, such as significant variation of the dark energy around the epoch when it comes to dominate or an extended, local `bubble' in which the Hubble parameter differs systematically from its large-scale average (as has been suggested by some in the literature).

Data Quality and Control of Systematics

Another critical rationale for the design of the SDSS SN survey is the desire to achieve high data quality and, thereby, improved control of systematics for precision SN distances, beyond what will be achieved at higher redshift in the next several years. By high data quality we mean: (1) a uniformly calibrated, well-characterized photometric system on a single instrument whose response is repeatedly monitored (capitalizing on the years of effort that have gone into understanding and calibrating the SDSS photometric system); (2) dense light-curve sampling in multiple optical bands, which should lead to reduced uncertainties in K-corrections and host-galaxy reddening; (3) spectroscopic follow-up at good signal-to-noise ratio for all high-quality SN Ia candidates and multi-epoch spectrophotometry for a significant fraction of the Ia sample, which should also help reduce K-correction uncertainties. Dense sampling, and therefore early SN detection, will be achieved by visiting each sky region every other night; it should allow improved modeling of the SN lightcurves (measurement of light-curve shape) and thus more accurate correction for the luminosity-decline rate relation. The lower redshift portion of the SDSS sample will have multiple epochs of spectroscopy, along with well-sampled NIR lightcurves from the Carnegie Supernova Project (CSP), providing a powerful basis for studying and controlling SN systematics and for exploring additional correlations that may further reduce the scatter in the Hubble diagram.

Evolution of the progenitor population remains one of the primary

systematic concerns for SN distance measurements at high redshift.
At z = 0.3, the lookback time in the LambdaCDM cosmology is about 3 Gyr; at z = 0.8, it is about 7.5 Gyr. Thus, systematic population evolution effects should be reduced in the SDSS sample compared to higher redshift surveys.

Since SNe probe cosmology by measuring relative distances between low and high redshift, another systematic concern with existing SN surveys is that they must combine heterogeneous data from different telescopes and differing photometric systems. While the full cosmological power of the SDSS SN survey will come from combining it with higher-redshift data, it will be unique in its ability to bridge low and intermediate redshifts, and thereby obtain cosmological constraints, in a single survey and system. (In this context, we note that the SDSS filter system is now being used at lower redshift by the CSP and at higher redshift by CFHTLS.)

Probing the dark energy with substantially increased precision has been identified by several national panels and interagency reports as a very high U.S. science priority; future projects collectively costing over 1B$ are being planned to address it. SNe Ia distances will be an important component of that program. In order to lay the groundwork for and enhance the utility of those projects, the SN Hubble diagram must be measured with greater precision and more complete redshift coverage in the coming years. This requires not only increased statistics (more SNe) but improved control of systematics. The SDSS SN survey is designed to address these needs by targeting an important, underexplored region of the Hubble diagram with a sample that, by virtue of its high photometric quality and extensive spectroscopic follow-up, will control (and explore) systematics at a high level.

Color Selection of Other Types of Supernovae

The SDSS SN survey will encompass the largest sky area by far with uniform, frequently sampled, 5-band photometric coverage. It will therefore have unique ability to discover, and select for spectroscopic follow-up, rare supernova types based on their colors, including, e.g., the type Ibc hypernovae thought to be associated with GRBs. A portion of the spectroscopic follow-up program will be devoted to this.
For the same reasons, this photometric dataset will be of substantial interest to those studying other time domain phenomena.

3) Recognizing that there are certain scheduling efficiencies built into the project as proposed, what would be your priorities among the 3 science areas proposed? How would these priorities be incorporated into the scheduling of observing time and the data processing and reduction?

Much of this question is answered in the answer to Question 1. Concerning priorities for processing and reduction, the spectroscopic processing is so fast that it is never an issue. Typically we have fully reduced spectroscopic data within two days of its acquisition. The processing of the imaging data takes about 10 days, which includes applying all of the calibrations. This is fast enough that we can keep up with the data flow, and use the object catalogs for subsequent spectroscopic plate drilling. Since the rate of data coming off the mountain will not change in SDSS-II for Legacy + SEGUE, we do not anticipate needing to prioritize the processing. We expect always to have a sufficient backlog of drilled plates for both surveys that the observing schedule is buffered from any decisions about what to process when.

The Supernova program is such that a string of clear nights will result in the processing at Fermilab lagging the rate of acquiring these data. However, the critical time-sensitive identification of transient objects will come from processing on the mountain, and the compute power for that operation is designed for a 24-hour turn-around. The corrected frames from the Fermilab processing will be available on the time scale of a month, since the processing can catch up during the bright run.

Near the end of the November dark run, we may undertake imaging for both the Supernova program and for SEGUE. In principle we would need to prioritize the processing, since SEGUE may need plates drilled for the next dark run. In this case we would give SEGUE priority.

4) Can you assign a cost to each of the science projects? Were the supernova search component dropped, what would be the savings? Could the observing efficiency on the other two projects be increased?

Most of the costs of undertaking SDSS-II depend on the combined survey duration, since we work with a dedicated staff throughout the year. The software development costs associated with Supernova are about 2.5 FTE-years or $175k. Individuals performing this work include scientists, post-docs and computer professionals. New computing hardware to support Supernova data processing and distribution will cost approximately $70k, and an additional $15k will be required to support travel and other expenses, for a total of $260k.

The savings from dropping Supernova can be estimated as follows. We make the calculation on the basis that the time required is that needed to make sure that Legacy and SEGUE are completed, namely three observing seasons, no matter what else is done. We will also assume that time not used for Supernova will be used for SEGUE rather than leave the telescope unused during periods when Legacy cannot be pursued. With these assumptions, the changes in costs incident upon dropping Supernova are as follows:

Changes in operating costs: $0.0M (no change in total time)

Changes in development costs: -$0.26M (SN s/w and h/w savings)

Changes in consumables: +$0.03M (115 extra SEGUE plates)

==

Net change in costs -$0.23M

Thus, we estimate that there would be a cost savings of $230k if Supernova were dropped. However, new partners who are contributing specifically to the overall project because of the existence of Supernova are providing us with $175 in additional cash. So, in the net, dropping Supernova would save the project only $55k.

Concerning the question about an increase in efficiency for Legacy and SEGUE: Legacy would not be helped by dropping the Supernova survey because its footprint is on the opposite side of the sky. For SEGUE there would be some gains because there would be fewer constraints, but SEGUE also needs to obtain spectra in the North Galactic Cap, and that would not be helped by dropping Supernova.

5) Will processing of SEGUE data at low-b require development of new processing techniques, and will the SEGUE data be statistically useful to achieve the scientific objectives? Accurate photometry at low-b appears to be substantially more challenging than NGP photometry. The size of the statistical sample in SEGUE is not much larger than existing samples. How is the budget justified in terms of the improvement in statistical sample? What is the statistically-achievable proper motion accuracy in SEGUE?

Low-b and crowded-field photometry

The SEGUE science goals are focused on gaining a large-scale understanding of the kinematics, chemistry and density distribution of the Milky Way's old stellar populations. By design, SEGUE requires no new software development for crowded-field, low-latitude photometry in order to achieve those goals, and our plan is to use the exact same pipeline code and infrastructure as is currently being used for the SDSS.

The current SDSS imaging survey cuts off in Galactic latitude at $|b| \sim 30$ deg, well before crowding in the plane becomes important. We have used both SDSS imaging scans from the main survey, together with test SEGUE scans, to evaluate the stellar density at which the photometry from the SDSS pipelines is no longer reliable.

From SEGUE test imaging scans that extend through the Galactic plane and which have been processed using the present pipelines, we find we can work down to |b|=8 deg at l=110 deg and to |b|=15 deg at l=50 deg, both representing a stellar density of approximately 50,000 stars per square degree to g = 23 (14 stars/sq arcmin). Crowding effects are due to the many faint objects below this limit and the extended PSF wings of brighter stars.

We have also verified that our spectroscopic target selection algorithms work down to these latitude limits in regions where the reddening (Schlegel et al., 1998) is E(B-V) < 0.3. We compute that 87% of the area which SEGUE plans to scan is above these |b| limits, which we vary appropriately as a function of longitude. Nearly all of the 200 SEGUE pointings are directed toward lines-of-sight that avoid these high stellar density or extreme reddening limits, and the associated photometry meets the requirement of 2% rms absolute and 2% in g-r.
These photometric errors are for objects with g < 21 in uncrowded fields. Magnitude-dependent systematic photometric errors of about 5% at g = 21 are present as the crowding density doubles, though frequently reddening corrections are a larger source of error, especially in the bluer filters (u,g,r).
The photometric sample reaches its 5 sigma limit at about g = 23.

The imaging camera will not be turned off as SEGUE scans through the plane of the Galaxy. These data will be made available, along with any current calibrations, as part of regular SEGUE data releases for interested researchers.

Several SEGUE partners are interested in pursuing independent analysis of crowded field data, using either existing (i.e. DAOPHOT) or new crowded field codes. This will not impact the development budget for SEGUE.

The SEGUE statistical sample that falls between the |b|=30 limit of current SDSS data and the low-latitude limits set by the pipelines will provide 2000 square degrees of new imaging data that are crucial to meet the SEGUE science goal of obtaining a global picture of the Galaxy's major structural components. One of the main science drivers for the set of low-latitude SEGUE imaging scans is to extend the photometric and spectroscopic sample of old, high angular momentum stars we associate with the thick disk to larger distances from the solar circle than the NGP-centered cone of the SDSS. At Galactic latitude b=10 (at l=110 deg), a main-sequence turnoff star at the nominal age and metallicity of the thick disk ($M_r \sim 4$) at the SEGUE r = 19 spectroscopic limit is located 1.7 kpc above the plane at a distance of 10 kpc in the plane. Within the limits set by the SDSS pipelines, SEGUE will sample a large range in Galactocentric distance near the plane and get the data necessary to build up a global picture of the thick disk and its relation to the stellar halo.

Statistical sample

We believe that the large volume of the Galaxy that SEGUE will survey, including precise multicolor optical photometry of millions of stars, combined with radial velocities and full spectral coverage at moderate resolution ($R \sim 1800$) of 240,000 objects specially targeted for Galactic structure studies, makes the SEGUE statistical sample unique among existing or proposed surveys which will complete within the next decade.

The 2MASS infrared imaging survey is all-sky, and is excellent at probing low-latitude structures photometrically. However, it is not deep, ($J < 16$, compared with $r < 23$ for the SDSS) and can only sparsely sample such features as the Sagittarius tidal stream by using relatively rare M-giants as tracers. It also has no built-in spectroscopic followup for key velocity information, and does not have the optical baseline colors, such as u-g and g-r, essential for precise photometric population separations and photometric parallax determinations.

The RAdial Velocity Experiment, RAVE, is a large, high-precision spectroscopic survey to examine the local structure of the Galaxy in detail. In its current phase using the 6dF facility at the 1.2m UK Schmidt telescope, which is part of the Anglo Australian Observatory, RAVE targets stars in the range $9<I<12$, $0.4<B-V<0.8$. This sample extends to distances of 1.5kpc from the sun. The RAVE velocity accuracy is specified to be better than 2 km/s and the spectra cover the Ca triplet region, from 8500 to 8750 Angstroms, at a resolving power of R=8000. A typical S/N ratio of 50 is obtained, allowing derivation of some stellar parameters, such as abundances and temperatures. The software pipelines for the stellar parameter determination are still under development.

As of December, 2005, RAVE has acquired 50,000 of the 150,000 stars planned for the phase of the project that is currently funded. RAVE aims to be a complete magnitude limited survey of stars in the solar neighborhood. A proposed follow-on project to extend the survey to magnitude V=16, including 25 million stars, has not been funded as yet. The current intention is to continue 6dF use, using dark-time for some deep pencil-beams, probing distances out to 10kpc and beyond with the BHB and giant branch stars in this sample. The majority of the survey would concentrate on a brighter limit, more complete survey. Being southern hemisphere, and concentrating on near-complete coverage of brighter stars, RAVE is designed to be complementary to SEGUE.

SEGUE will use the high-efficiency SDSS spectrographs to get radial velocities for 240,000 stars at significantly fainter magnitudes, $14.5 < g < 20$, than RAVE. SEGUE will sample a much larger volume of the Galaxy, with an emphasis on large distances toward the Galactic anticenter, regions where substructure retains its coherence longer than a few Gyrs (Helmi and White, 1999) because streams experience less of the disruptive effects of the Galactic disk and have velocity dispersions well-matched to SEGUE's 5-10 km/s radial velocity accuracy. In contrast, RAVE is designed to unravel the merger history of the stars in the nearby Galaxy.

The most accurate large-scale astrometric survey is the Hipparcos catalog, but its volume is small and limited to the very nearby solar neighborhood, as are currently available high-resolution spectroscopic samples such as Edvardsson et al. (1993 AA 275, 101) and the newly completed Geneva-Copenhagen survey (Nordstrom et al., 2004 A&A 418, 989) which extend to about V=9. When it is released, UCAC (Zacharia, N. 2004 AN 325, 631) will extend that volume to stars with $V < 16$.

At further distances, two surveys which focus on distant stars are the Spaghetti survey (Morrison et al. 2000 AJ 119, 2254) for K giants in the halo and the Century Survey Galactic Halo Project (Brown et al. 2003 AJ 126, 362) of bright distant blue stars. Both of these surveys have made valuable contributions toward techniques for reliably identifying halo objects. Their sample sizes are considerably smaller than the proposed SEGUE sample.

In the next decade, the space missions GAIA and SIM will provide accurate astrometry to much larger distances, opening up a larger volume of the Galaxy to full, 6D phase-space studies. SIM is a NASA mission scheduled for launch in 2010. Over five years, it will achieve an astrometric accuracy of up to 4 microarcseconds for a set of pre-selected targets. The SIM Galactic Structure Key Project will observe ~6500 stars in the Galaxy and its nearby companions. The goal is to investigate the mass and mass distribution of the Galaxy from the central regions to the outer halo. GAIA is an ESA mission scheduled for launch in 2011. Over five years, it will obtain astrometry to 10 microarcsec at V=15 and 20 milliarcsec at V=20 for all the stars in the sky to V<20, and radial velocities to 1-10 km/s at V = 17-18. Thus radial velocity follow-up for the faint end of the GAIA dataset will be necessary in order to get full space velocities for the majority of Galactic halo stars in the GAIA sample. The SEGUE spectroscopic survey will make the first major contribution to that effort, and the data will be available well before the matching GAIA data are taken. SEGUE will be finished before either of these two missions begin operations, and the data and Galactic science results from SEGUE will provide input to the design and observing strategy of both projects.

The SEGUE spectroscopic targets are chosen to sample intervals in log(distance) with ~100 radial velocities per interval. The spectroscopic sample covers the Galaxy in a grid of 200 sight lines with a 10-20 degree typical spacing so that all major components (thick disk, thin disk and halo) and streams (Sagittarius, Monoceros/Canis Major, Tri-And, other unknown), which all have dimensions greater than this scale, will be sufficiently sampled. SEGUE samples the 1.5 < d < 6 kpc thick disk/spheroid transition region with a uniformly selected sample of G dwarfs and it probes the halo and tidal streams with spectroscopy of F, BHB and K giant stars at distances from 5 < d < 100 kpc. The SEGUE spectra cover a broad spectral range, 3800A-9200A, which provides leverage for giant/dwarf separation and indicators for metallicity and effective temperature.

The 100 spectra per pointing per distance interval of SEGUE allows detection of deviations from non-Gaussianity in the histogram of velocities, essential to mapping out global picture of all major Galactic components, including streams, known and unknown.

The large area and coverage of the distant Galaxy, especially toward the anticenter, make SEGUE uniquely suited to taking a census of substructure in the outer Galaxy.
Surveys like RAVE, SIM and GAIA will search the nearby volume, where precision kinematic data is necessary to find substructure that has phase-mixed through interaction with the gravitational potential of the disk. SEGUE will take its census further out where debris streams are more coherent but dispersed over smaller

angles on the sky, requiring large survey volumes. When added to the local measures of the Galactic accretion history, we will have a full accounting of the substructures that have been incorporated into the Milky Way. Comparison of the observed and predicted merger history of the Galaxy will directly address issues of CDM substructure on small scales, and is a fundamental test for models of galaxy formation.

Similarly, the study of the thick disk requires not only SEGUE's large statistical sample, but also the large volume of the Galaxy that SEGUE will probe. The formation of the thick disk is probably linked to that of the thin disk, which contains most of the baryonic mass in the Galaxy. This close link makes the thick disk an especially important piece of the Galaxy's past. Its large vertical velocity dispersion should make the thick disk immune to most mechanisms for radial mixing that operate in the thin disk. A sample limited to radii near the solar circle will not be as sensitive as SEGUE to radial gradients in chemical abundance or kinematics, nor to any coherent substructure in the outer disk, all of which may be important clues to the formation of this fossil component of the Galaxy.

In addition to the many Galactic structure studies that will be done within SEGUE, the large, homogeneous, well-characterized survey released to the community will be a resource that can fuel other, follow-on investigations. For example, the photometry and moderate-resolution spectroscopy will provide the raw material for finding extremely low-metallicity candidate stars ([Fe/H] < -3) for followup by on 8-10m class telescopes and their next generation instruments.

Thus, we believe that the large statistical sample and large volume probed by SEGUE will be unique, unmatched by any other survey for at least a decade, and that these characteristics of the SEGUE dataset are absolutely essential for addressing some of the most interesting and timely questions about the structure and evolution of the Galaxy.

Proper motion accuracy

The astrometric accuracy of the SDSS imaging data is approximately 50 mas (milliarcseconds) per coordinate. Munn et al. (2004 AJ 127, 3034) have used the large-area object catalogs from the SDSS to recalibrate the USNOB (Monet et al. 2003 AJ 125, 984) proper motion survey. Using a 50 year baseline, they have obtained final errors on the proper motions using this method, as measured by the SDSS sample of spectroscopically confirmed QSOs, of 3.5 mas/year (one sigma) in ra and dec for individual objects with r < 19, where the catalog is highly complete. For objects with 19 < r < 20, completeness drops to about 80% and the accuracy decreases to 4.5 mas/year.

To give an example for a three-sigma motion of 10.5 mas/year, a star at 1 kpc (e.g. a K dwarf with M_r = 7.0 and r = 17) would be detected with a transverse velocity of 17 km/s, sufficient to explore the velocity ellipsoid in the transition region from the thin to the thick disk with the large projected SEGUE sample.

Similarly, for a star at 2 kpc (a G dwarf with M_r = 5.0 and r = 16.5), we will detect motions corresponding to 33 km/s. Using the large sample of G and slightly brighter F stars, SEGUE can distinguish rotating halo components from pressure supported ones.

Beyond 2kpc, SEGUE will transition to primarily using radial velocities and distance estimates based on photometric and spectroscopic indicators. However, proper motions will still be useful in a statistical sense when averaged over large numbers of similar stars toward similar directions on the sky. Additionally, the degree to which we can separate distant K giant candidates from foreground K dwarfs is significantly improved with the available astrometry.

6. Additional budget detail is needed for the reviewers to adequately assess the magnitude of the request. Please make a clear presentation of what hardware, software, etc. for the SDSS project can be used for the extension and what must be acquired and developed.

All of the hardware and software developed for and used in SDSS-I operations will be used for SDSS-II, since the Legacy survey is a continuation of current operations. At the observatory, this includes the 2.5m telescope, 0.5m Photometric Telescope (PT), imaging camera, two spectrographs, PT camera, nine fiber cartridges, and the many ancillary support systems at APO. It also includes the software used to control the telescopes and instruments and the data acquisition system used to acquire data from the instruments and write it to tape for subsequent data processing.

At Fermilab, where the SDSS data are processed, calibrated and prepared for distribution, the entire existing infrastructure will continue to be used for SDSS-II. This includes a large array of computing hardware to store and process data, the imaging data reduction pipeline (PHOTO), the spectroscopic reduction pipelines (idlspec2D and Spectro-1D), the photometric telescope data reduction pipeline (MTPipe), the quality assurance tools developed to assess and verify data quality, target selection software (Target) used to select objects for follow-up spectroscopy and design spectroscopic plug plates, and numerous other infrastructure tools that have been put into place over the past four years to make the data processing operation run efficiently and effectively.

At Princeton, where spectroscopic data for the SEGUE survey will be processed and calibrated, and where the photometric calibration work will be centered, there exists a photo-op and spectro robot system. This system has been used during SDSS-I operations to independently reduce and calibrate SDSS-I imaging and spectroscopic data and the existing infrastructure will be used for SDSS-II. Thus, while there are detailed development plans associated with the photometric calibration effort and SEGUE spectroscopic data reduction, for the most part these are relatively minor revisions to a working photo-op and spectro robot system.

The available infrastructure also includes the software and hardware used to serve processed data to the collaboration and general public. This includes the Data Archive Server (DAS), Catalog Archive Server (CAS), and the SkyServer and CasJobs user interfaces to the CAS.

At the University of Washington, the tooling currently used to fabricate plug plates will continue to be used for SDSS-II. This includes a specially-designed drilling fixture that forms plug plates into the required shape for drilling, the large vertical drilling machine that has been modified with a special temperature control system to maintain a constant machine temperature throughout the drilling operation, the fixturing and coordinate measuring machine (CMM) used in our QA process to verify proper hole diameters and positions, and the software used to operate the drilling machine, perform the QA measurements, and record the results.

The new hardware and software required for SDSS-II is best summarized by project.

For Legacy, the only new hardware and software required is that associated with the DA upgrade. New hardware for the DA includes the following:

- Nineteen Motorola PowerPC boards with SCSI option, for each crate (13 for the DA system, 4 for a test stand, plus 2 spares);
- Two high-end dual processor PCs to replace *sdsshost* and *sdssmth* (4 GB, ~2TB disk);
- Five 250GB IDE drives and two 8-bay disk chassis';
- Eighteen DLT tape drives to replace current drives which are nearing the end of their expected life.
- One Motorola PowerPC board, one transition module, and a Linux host with disks, to upgrade the DA test stand at Fermilab;
- VxWorks license for the PowerPC platform.

Software development includes revisions to existing code so that it runs on the new PowerPC platform and emulates all of the functionality of the current system. The scope of work associated with the DA upgrade can be found at http://tdserver1.fnal.gov/sdss/SDSS-II/WBS/daUpgradeWBS.pdf .

We intend to complete the DA upgrade development work during the spring and summer of 2005. System installation and commissioning at APO is scheduled for August 2005, with the system fully tested and ready for use in September 2005, in time for the fall 2005 observing season.

For SEGUE, software-related development is summarized as follows:

- Develop stellar atmosphere parameter code in order to refine the estimation of stellar temperatures, metallicities, and gravities;
- Finish the development of the target selection code by refining the algorithms for the eleven targeted object categories, and validate the code based on empirical data;
- Refine existing data quality assessment tools to monitor and verify SEGUE data quality during operations;
- Revise the spectroscopic pipeline (specBS) to incorporate the most-recent ELODIE star templates to improve fitting algorithms for pipeline radial velocities and star-typing;
- Characterize errors in metallicity, gravity, T_{eff} and radial velocity as a function of magnitude, stellar type and observed signal/noise;
- Revise the spectroscopic pipeline (idlspec2D) to output the sky spectra in addition to the improved sky-subtracted spectra, and to include fully calibrated individual 15-minute exposures;
- Develop code to package new spectro parameters into uniform flat file format for collaboration use;
- Modify existing DAS and CAS data distribution software to accommodate SEGUE data.

New hardware will be required to process SEGUE spectroscopic data and to serve SEGUE data to the collaboration and general public. For spectro data processing, a dual-processor high-end PC, outfitted with sufficient memory and disk space and running Linux, will be required. Additional disk space will be added over time as the volume of data increases. The estimated cost of this hardware is $30K. For data distribution, three dual-processor database servers, each with 6 TB of SATA disks configured in RAID5 arrays, will be required to load and serve data to the collaboration and public. Estimated cost of this hardware is $40K. The scope of work associated with the SEGUE project can be found at http://tdserver1.fnal.gov/sdss/SDSS-II/WBS/segueWBS.pdf .

Work is underway on many aspects of the SEGUE development task list. Our intent is to have the work associated with SEGUE target selection completed by July 1, so that spectroscopic plug plates can be designed and fabricated in time for the fall 2005 observing season. Work in other areas will continue into the fall observing season; none of this work will impede the start of SEGUE observing this fall.

For the Supernova (SN) project, software and software-related work involves the following:

- Installing a data processing system at Apache Point Observatory (APO) to perform next-day reductions on newly acquired data;
- Modifying the Frames Subtraction pipeline to implement an improved remapping algorithm between search and template frames; to improve subtracted image noise characterization, to implement better re-sampling methods; and other improvements.
- Further developing software scripts to automate data processing;
- Further develop a Supernova Candidates Database and associated web interface;
- Implement I-band frame subtraction for all SN candidates;
- Refine spectroscopic target selection code by developing color-color and color-magnitude pre-selection criteria and implementing real-time light-curve fitting;
- Further developing frameworks and tools for follow-up observations;
- Further developing pipelines for off-mountain data analysis, in particular final SN photometry measurements;
- Developing databases to serve SN data, and repeat imaging data and/or catalogs, to the collaboration and general public.

New hardware will be required for the SN data processing system at APO. Benchmark tests are currently underway to determine the precise type and quantity of hardware required to process newly acquired SN data through *gri* frame subtraction in 24 hours or less. At present, we estimate that the system will require 8 to 10 dual processor CPUs, each running at 2 GHz and configured with 1 GB RAM and 1.5TB of SATA drives in a RAID array. The estimated cost of a compute cluster with 8 machines is $40K. The estimated cost of database servers to host and serve SN data to the collaboration and general public is $30K. Also, the addition of the SN compute cluster at APO will increase the heat load in the existing computer room. The increased heat load will be partially offset by the reduction in heat load that we will realize through the DA upgrade; replacing old computers with fewer, more efficient units will result in a smaller heat load from the DA system. Once the machines for the SN cluster are selected, taking heat load considerations into account, we will complete a careful analysis of the total heat load on the computer room and upgrade the cooling capacity only if absolutely necessary. The scope of work associated with the Supernova project can be found at http://tdserver1.fnal.gov/sdss/SDSS-II/WBS/supernovaWBS.pdf .

Our intent is to have all of the initial software development work for the Supernova project complete by June 1, 2005, with the exception of the SN Photometry Pipeline. We anticipate completing the initial version of this pipeline by the end of June. We also intend to have the new Supernova compute cluster in place at APO in mid-July, with installation and commissioning finished by early August. This will put all pieces in place for the Supernova project by the start of the fall observing season in September.

SDSS photometric calibration is now good to of order 2% rms (although with non-Gaussian tails); using scan overlaps and a series of perpendicular 'Apache Wheel' scans to tie the photometry together promises to improve this to of order 1%. The work on the photometric calibration effort is purely analysis and software related; no new hardware is required to support this work. Development of software to reduce Apache Wheel data is well along, and the Apache Wheel data collected during SDSS-I is now being reduced. While further iterations on the code will likely be required, we anticipate that the software will be ready (as a prototype) by the time SDSS-II is running.

The use of the overlaps and Apache Wheel data in the photometric calibration effort has happened in a software package called 'photo-op', using code called 'ubercalibration'. The ubercalibration results also result in improved flat-fields, based on a series of oblique imaging cross-scans. At Princeton, where spectroscopic data for the SEGUE survey will be processed and calibrated, and where the photometric calibration work will be centered, there exist both a photo-op and spectro robot system. This system has been used during SDSS-I operations to independently reduce and calibrate SDSS-I imaging and spectroscopic data and the existing infrastructure will be used for SDSS-II. Thus, while there are detailed development plans associated with the photometric calibration effort and SEGUE spectroscopic data reduction, for the most part these are relatively minor revisions to a working photo-op and spectro robot system. There is work to feed the results into the data processing factory at Fermilab.

We expect that the person running the photo-op pipeline and applying the ubercalibration code will also run the specBS pipeline to process the SEGUE spectroscopic data, and integrate stellar parameters code into either SpecBS or (more conveniently) a post-pipeline to be run on the classified outputs of SpecBS.

In detail, the tasks associated with the photometric calibration effort include:

- Integrating the ubercalibration code into photo-op and providing testing outputs;
- Testing the output calibrations for systematic effects and statistically significant spatial structure;
- Designing, and writing code to make, the FITS binary tables to use for post-calibration of the Fermilab outputs;
- Writing code to extract the flat fields from the photo-op code and deliver them to Fermilab in a timely fashion after cross-scan calibration;
- Analyzing white dwarf spectra and other spectrophotometric data to provide the best possible calibration of the SDSS photometry onto an AB system;
- Run the both the photo-op and spectro robots. This is primarily a task of keeping track of disk space, understanding changes in mountain procedures, and getting the data into the data processing system at Princeton. This is not a major task, but requires care and attention to detail.

The budgeted level of effort associated with the calibration effort is 0.25 FTEs in 2005, 0.50 FTEs in 2006, and 0.75 FTEs in 2007 and 2008.

With regard to budget detail, the SDSS-II cost estimate is organized by Work Breakdown Structure (WBS), as follows:

1.0 Survey management……………………………..…$1,629K

2.0 Survey operations
- 2.1. Observing Systems………… $ 2,552K
- 2.2. Observatory operations……....$ 5,174K
- 2.3. Data processing………………$ 2,390K
- 2.4. Data distribution…………..….$ 1,451K
- 2.5. ARC support for operations.…$ 207K
    - Survey Operations sub-total…………..……$ 11,774K

3.0 New Development
- 3.1. SEGUE development……….....$ 369K
- 3.2. Supernova development………..$ 258K
- 3.3. Photometric calibration…………$ 146K
- 3.4. Data acquisition system upgrade...$ 241K
    - New development sub-total…………………..$ 1,013K

4.0. ARC Corporate Support …………………………..$ 179K
5.0. Public Outreach …………………………………..$ 0K
6.0. Management Reserve……………………………..$ 305K

TOTAL……………………………………………..$ 14,900K

The organization of costs is very similar to that included as Exhibit 1 in the NSF proposal. The one area that is organized differently is Section 3; we modified the WBS so that development work and costs are organized by development project, rather than by deliverable, at level 2 in the WBS. The deliverables associated with new development work are now organized by sub-project and captured at level 3 and lower in the WBS. The SDSS-II WBS can be found online at http://tdserver1.fnal.gov/sdss/SDSS-II/WBS/SDSS-II-WBS.pdf .

The process of reorganizing the WBS resulted in a shifting of costs between categories. While the total cost of the project remains at $14.9 million, the distribution of costs between operations and development is different than that shown in Exhibit 1 of the proposal. The redistribution of costs is due in part to the WBS reorganization and in part to a reallocation of resources based on a better understanding of the development work scope.

Sections 1, 2, and 4 of the WBS constitute the core aspects of running the survey; the cost estimate for these sections is based on the experience gained during the past four years of SDSS-I operations. In preparing the cost estimate for SDSS-II, each cost element in each section was carefully scrutinized. We reviewed staffing needs in each section; as a result, the size of the staff supporting SDSS-II operations will be slightly less than that employed for SDSS-I. We also reviewed the budget for travel, materials, supplies and observatory operating expenses, and made reductions where possible without impacting safety, equipment protection, data quality, or operating efficiency.

As a point of reference, the actual annual operating cost for SDSS-I, for the years 2001-2004, was $4.86 million per year, on average. The estimated average annual cost going forward into SDSS-II operations (sections 1,2,4, and 5) is $4.54 million /year.

Section 3 of the WBS captures the work and budget associated with new development. The development budget is based on the level of support required to develop the software and systems necessary to achieve the science objectives of the survey. The development budget can be broken down as follows:

** The SEGUE development budget provides 4.0 FTE-years of salary support to develop the data reduction and distribution software for the SEGUE program. The SEGUE budget also provides $70K for data processing and distribution hardware and an additional $17K for travel and other miscellaneous expenses.

** The Supernova development budget provides 2.5 FTE-years of salary support to develop the hardware and software systems to process and distribute Supernova data. The Supernova budget also provides $70K for data processing and distribution hardware, and an additional $15K for travel and miscellaneous expenses.

** The Photometric Calibration budget provides 2.25 FTE-years of salary support to finish the development and testing of the calibration software, compare results to externally calibrated data, and develop the means for integrating the calibration into the production data processing operation. The budget also provides an additional $4K for travel and miscellaneous expenses associated with this work.

** The data acquisition system (DA) upgrade is necessary to address obsolescence and reliability issues present in the current system. Some critical components of the DA system are obsolete and no longer available. In addition, there are hardware problems in the existing system with the potential to seriously impact observing operations. The DA upgrade budget provides 1.15 FTE-years of salary support for computer professionals to upgrade system hardware with current technology and port existing code onto the new platform. The budget also provides $105K for hardware replacements and $10K for travel to install and commission the improved system.

7) Will SDSS-II incorporate "outsiders" on the science teams?

Yes. We have successful mechanisms to include others for the SDSS-I, and we will continue the same policies in the SDSS-II. These policies are available at http://www.sdss.org/policies/index.html. One of the most relevant policies relates to a status we call "External Collaborators." If an external person can provide special expertise, code, telescope time, etc., that enhances the science we can do with SDSS, that person can be nominated for this status. We have a standard review process to check that there are no conflicting interests. If approved, the person is granted access to SDSS data to undertake a specific project. These individuals are regarded as members of the SDSS Collaboration for the duration of their project - for example, they may attend Collaboration meetings, and we have included some as Invited Speakers in the Special Sessions that we have organized at AAS meetings. They can participate in scientific and technical discussions via the Working Groups and other mechanisms (e.g. mailing lists). Many are responsible for the highest-profile science, such as the galaxy power spectrum papers. Fifteen of the approximately 85 refereed papers published in 2004 had External Collaborators as either the first or second authors. To give another sense for how active this program is, we list below the External Collaborators from just the year 2004. (By definition these projects will terminate at the end of the SDSS-I, but it is reasonable to expect that many of these same people will reapply for similar projects under SDSS-II.)

Cullen Blake (grad student, Harvard)
Nathan De Lee (grad student, Michigan State)
Young Sun Lee (grad student, Michigan State)
Horace Smith (Michigan State)
Ignacio Ferreras (Zurich)
Joe Hennawi (Berkeley)
Tim Beers (Michigan State)
Carlos Allende Prieto (Texas)
Ron Wilhelm (Texas Tech)
John Norris (ANU)
Eva Grebel (Zurich)
Alexei Kniazev (ESO)
Mike Gladders (Carnegie)
Paul Smith (Steward)
Gary Schmidt (Steward)
Chris Howk (UCSD)
Phillipp Richter (U. Bonn)
Varsha Kulkarni (South Carolina)
I. Neill Reid (STScI)
Alejandro Clocchiatti (Universidad Catolica)
Martin White (Berkeley)
Alex Gray (CMU)
Adam Myers (Illinois)
Masafumi Yagi (NAOJ)
Karen Vanlandingham (Steward)
Pablo Fosalba (Hawaii)
Pat Hall (York)
Russet McMillan (Apache Point Observatory)
Bill Ketzeback (Apache Point Observatory)
Jack Dembicky (Apache Point Observatory)
Jun Pan (Hawaii)

Stephane Colombi (IAP)
Fred Hamann (Florida)
Paola Rodriguez (grad student, U. Florida)
Dimitri Pourbaix (Belgium)
Frank Bertoldi (MPIfR)
Pierre Cox (IAS, France)
Chris Carilli (NRAO)
Fabian Walter (MPIA)
George Rieke (Steward)
Dean Hines (Steward)
Emeric Le Floch (Steward)
Marianne Vestergaard (Steward)
Maarten Schmidt (Caltech)
Richard Green (NOAO)

Besides External Collaborators, there are other ways to include people not at SDSS institutions, and we will extended those respective policies forward to SDSS-II as well. These instances relate to younger scientists who move from an SDSS institution elsewhere during the course of undertaking a project, and including (as full Participants) a small number of people who are part of the operations team.

8) Several reviewers noted the difficulty of use of the SDSS database to the average astronomer. Have you given any thought to improving the interface, and what changes could you make to make the present and future database easier to use?

We are committed to providing a public window to the SDSS data that is as useful as we can make it. The SDSS data are rich in information, and as a consequence the data and their interrelationships are complex and do require a learning curve for the effective use of the interfaces. Moreover, we provide more than one interface, and these options may underly some of the concerns about ease of use. But with extra options comes the benefit of extra power: once the basic features of the SDSS data and interfaces are understood, the astronomer is rewarded with an exceptional research tool.

It is possible that some of the perception of difficulty comes from the query interface, which requires familiarity with SQL. We are finding that SQL usage is doubling approximately every 3.4 months, as opposed to the web usage in general, which doubles in approximately 15.4 months (and is now exceeding 2 million hits per month). This analysis suggests that astronomers are willing to learn SQL since it offers a more powerful way to access the data than the simple forms. We have developed detailed tutorials and help pages to facilitate using the SQL query tool, including many sample queries. However, recognizing that many queries are simple ones like obtaining a finding chart or identifying objects near a given position, we also offer several straightforward form interfaces that require no knowledge of SQL. Each of these interfaces displays the full SQL query generated upon submitting the form, as an additional aid to learning SQL.

We have been considering ways to make the Data Archive Server easier to use and faster, for example by setting up a small database that would speed up the processing of requests for large sets of files, and that would allow searching for fpC files by Right Ascension and declination. Some of the Catalog Archive Server interfaces could be combined, thus simplifying things. For example, the spectro and upload crossID functions could be subsumed into the SQS and IQS form, where there is some partial overlap already.

The project has resources for making upgrades of this scale. We are eager to receive specific suggestions for improvements, and we will respond whenever it is practical to do so. With each Data Release, we review and upgrade the documentation, so there is an evolution via these improvements. We maintain a helpdesk to answer questions from users (one or two such questions per day), and when bugs are identified, they are fixed promptly.

9) What are your plans for long-term data management and curation?

The archives are naturally quantized in the form of (cumulative) data releases, the last of which will be DR8, scheduled for public release in November 2008. The DR8 database will be maintained at Fermilab, which will serve as the interim steward for the SDSS-II archive. Data distribution activities subsequent to DR8 will be related to maintaining the SDSS-II website (www.sdss.org), serving up the SDSS-II data archive, and providing helpdesk support. It is anticipated that Fermilab will provide these services until a long-term steward can be identified and the archive successfully transferred. By 2008, it is highly likely that several mirror sites will exist.

The project itself will be dissolved in 2008, and we cannot unilaterally commit other organizations for indefinite responsibilities. Recognizing that Fermilab may be a less natural home in the long run for an astronomy archive than, say, one of the NASA data centers, we made a presentation to NASA's Science Archives Working Group in October 2004. Our premise was that the SDSS-I and SDSS-II archives were important for space-based astronomy and would continue to be so in the long run. We argued that the expertise at the NASA data centers could be brought to bear, and provide NASA with an opportunity to support SDSS-II. The report of the SAWG is a supplied document. The key paragraph states:

"The SAWG sees the SDSS archive as an important resource for astronomical research both as a standalone data set and as part of the VO. We recognize that the SDSS archive is a testbed for the interagency VO efforts. The data and data base are of outstanding quality and value to the science community, so a plan should be developed to ensure the long-term curation of the SDSS archive. Since the data are ground-based and have been produced with significant support from NSF and the DOE, it would be appropriate for one or both of these organizations to provide the long-term support for the data set, with advisory input from NASA. A cooperative effort between agencies is not only desirable, but necessary for future endeavors (e.g., a VO) and cooperative efforts can be organized through cross-agency committees such as the AAAC."

Given this as background, our plan is as follows. Well before July 2008, a Memorandum of

Understanding will be drafted between the Astrophysical Research Consortium and Fermilab that provides the detailed expectations and terms for data management and curation of the SDSS-II archives at Fermilab, and the consequent resources needed to do the work. Fermilab will supply some of these resources, consistent with its contributions to other physics experiments, and consistent with its continuing support for astrophysics at Fermilab. However, Fermilab expects a partnership with other entities and agencies to supply the remaining resources, for example via a proposal from ARC to one of the agencies.

10) Does the current SDSS pipeline support rapid data releases which will be important to the SN program?

Yes. We have demonstrated that triggers can be generated in a timely fashion at the mountain, and these will be announced via IAU circulars. Additionally, all data will be processed at FNAL, and we have shown that we can process these with the full pipelines within one month of data collection. We plan to provide the data files (object catalogs and corrected frames) on that time scale.

11) Discuss progress made in any or all of these projects since the submission of the proposal.

The presentations by the Project Team at the 17 February review will include detailed progress reports which will be made available to the panel.